



TITLE:

欠損のある行列に対する上下制限 約付き低ランク近似 (数理最適化の 発展: モデル化とアルゴリズム)

AUTHOR(S):

達川, 愛美; 田中, 未来

CITATION:

達川, 愛美 ...[et al]. 欠損のある行列に対する上下制限約付き低ランク近似 (数理最適化の発展: モデル化とアルゴリズム). 数理解析研究所講究録 2018, 2069: 23-33

ISSUE DATE:

2018-04

URL:

<http://hdl.handle.net/2433/241963>

RIGHT:

欠損のある行列に対する上下制限約付き低ランク近似

達川愛美*

田中未来†

概要

本論文では、欠損のある行列に対する新たな低ランク近似モデルを提案する。提案モデルは上下制限約を含むものであり、これは協調フィルタリングへ応用することができる。この問題の厳密解を求めることは NP 困難だが、本論文では停留点へ収束する実用的なアルゴリズムを提案する。提案するアルゴリズムは交互最小化に基づくものであり、緩い仮定の下で停留点に収束する。

1 はじめに

1.1 背景

行列の低ランク近似は特徴量抽出によく用いられている。この手法では、高次元のデータを低次元の空間に埋め込むことができる。これは、高次元空間におけるデータがなんらかの関係をもつ場合、低次元の構造で表現できることが多いためである。特徴量抽出を用いることにより、データにおける潜在的な特徴を見出すことができる。これは、欠損のあるデータにおいてその欠損値を補完したり、ノイズのあるデータからノイズを除去したりするときに、その精度を高めることに役立つ。欠損のあるデータ行列を欠損のない低ランク行列に近似する応用例を以下に 2 つ示す。

1.1.1 欠損のある場合における主成分分析

主成分分析 (PCA) は特徴量抽出に用いられる古典的な手法である。これは高次元のデータを低次元空間に埋め込む手法である。低次元空間に埋め込まれた成分は主成分と呼ばれる。欠損のある入力データを扱う際には非線形なモデルが用いられるが、このことは過学習を引き起こしたり、局所解に収束したりする原因となる。Tipping and Bishop (1999) は PCA の定式化において確率モデルを導入した。これを確率的主成分分析 (PPCA) という。この手法では、欠損があるデータに対しても良い基底が求まることが知られている (Ilin and Raiko, 2010)。PPCA は EM アルゴリズムを用いて解かれることが多い。

* 東京工業大学 工学院 経営工学コース, ☎ 152-8552 東京都目黒区大岡山 2-12-1, tatsukawa.m.aa@m.titech.ac.jp

† 統計数理研究所 数理・推論研究系, ☎ 190-8562 東京都立川市緑町 10-3, mirai@ism.ac.jp

1.1.2 協調フィルタリング

行列の低ランク近似は、iTunes, Amazon, Netflix などで見られる推薦システムに用いられている。これらのサービスでは、ユーザの好みにあった音楽、本、映画などが推薦される。ユーザは聴いたことのある音楽、買ったことのある本、観たことのある映画などに対して評価をつける。そしてその評価に基づき、ユーザの好みそうな、あるいはユーザにとって目新しい音楽や本といったアイテムが推薦される。

推薦システムの例として、次のような状況を考える： m 人のユーザと n 個のアイテムがあり、各ユーザはアイテムのうちのいくつかに対して 5 段階で評価するとする。1 が最も低い点数であり、好んだアイテムほど高い点数をつける。各ユーザの各アイテムに対する評価値は $m \times n$ の行列で表現することができる。以下に $m = 3, n = 5$ のときの例を示す：

$$V = \begin{matrix} & \begin{matrix} a & b & c & d & e \end{matrix} \\ \begin{matrix} A \\ B \\ C \end{matrix} & \begin{pmatrix} 2 & 5 & ? & 4 & 1 \\ 1 & ? & 1 & 3 & 2 \\ ? & 1 & 4 & ? & 5 \end{pmatrix} \end{matrix}. \quad (1)$$

行列 V の各行はユーザ、各列はアイテムを表す。例えば、ユーザ A はアイテム b に対して 5 点をつけ、アイテム c には評価を与えていないことを表す。このように、記号 “?” は評価の欠損値を表すこととする。式 (1) にある行列からは、ユーザ A とユーザ B は好み似ているように読み取れる。これは、ユーザ A が高い評価をつけたアイテムはユーザ B も好むということを意味する。具体的には、アイテム b に対してユーザ A は 5 点をつけているが、ユーザ B も同様に高い評価をつける可能性があるということである。もしこの行列の欠損を埋めることができれば、あるユーザが評価していないアイテムに対し、そのアイテムをどれくらい好むかを予測することができる。この予測の手法は推薦システムに用いることが可能である。

推薦システムに用いられる手法は協調フィルタリングという。協調フィルタリングのアルゴリズムは主にメモリベース型、モデルベース型、ハイブリッド型の 3 つのカテゴリに分類される (Su and Khoshgoftaar, 2009)。メモリベース型のアルゴリズムでは、ユーザの嗜好を予測するために、ユーザ間、あるいはアイテム間の類似度を計算する。モデルベース型では、モデルを学習して予測を行なう。ハイブリッド型のアルゴリズムはいくつかの協調フィルタリングのアルゴリズムを組み合わせたものである。

メモリベース型では、ユーザ間やアイテム間の類似度を用いて予測を行なう。この類似度には、ベクトルコサイン類似度や Pearson の相関係数がよく用いられる。しかしながら、データに欠損が多い場合には、ユーザ間の類似度を計算することが難しい。実際のところ、アイテム数はユーザ数よりもはるかに多く、各ユーザが評価するアイテムはごく少数に限られる。このため、多くのアイテムはごくわずかなユーザによってのみ評価され、その他のユーザからの評価は欠損する。

メモリベース型アルゴリズムにおける上記のような欠点を克服するため、モデルベース型のアルゴリズムが研究されている。この手法ではデータマイニングや機械学習で用いられるアルゴリズムを用いる。モデルベース型のアルゴリズムにおいて用いられる手法の 1 つは、主成分分析や特異値

分解といった次元削減の手法である。これは上述の通り、高次元空間において関係をもつデータは低次元の空間に埋め込むことができることが多いため、高次元のデータであっても低次元の表現をもつと考えられているためである。

1.2 関連研究

行列の低ランク補完や低ランク近似は以前から研究されている。この節ではそれらに関する論文をいくつか端的に紹介する。

考えたいことは欠損のある行列 $V \in (\mathbb{R} \cup \{?\})^{m \times n}$ を補完することである。ここで記号 “?” は、この記号に対応する行列の要素が欠損していることを表す。つまり、 $V_{ij} = ?$ は V_{ij} が欠損していることを表す。Candés and Recht (2009) は行列 V を補完するために V のランクを最小化することを提案した。彼らが考えたのは以下の問題である：

$$\begin{aligned} & \text{minimize} \quad \text{rank}(X) \\ & \text{subject to} \quad X_{ij} = V_{ij} \quad ((i, j) \in \Omega). \end{aligned} \quad (2)$$

ここで、決定変数 $X \in \mathbb{R}^{m \times n}$ は V が補完された行列である。また Ω は添え字の集合であり、 $\Omega = \{(i, j) : V_{ij} \in \mathbb{R}\}$ とする。問題 (2) は、 l_0 ノルム最小化を本質的に含むため、NP 困難である。この困難さのある意味で引き起こしていると考えられるランク関数の非凸性や不連続性を取り除くために、彼らは問題 (2) の凸緩和として核ノルム最小化を導入した。核ノルム最小化は半正定値計画問題 (SDP) に帰着させることができる。内点法を含め、SDP を解くための効率的なアルゴリズムや高性能のソフトウェアは多数存在する。しかし、SDP を解くためにかかる計算時間は問題例の規模が大きくなると急激に増加するため、協調フィルタリングに由来するような大規模な問題例には適さない。

Olsson and Oskarsson (2009); Gillis and Glineur (2011) は、 V の補完を行なうため、以下の問題を考えた：

$$\begin{aligned} & \text{minimize} \quad \sum_{i=1}^m \sum_{j=1}^n W_{ij} (X_{ij} - V_{ij})^2 \\ & \text{subject to} \quad \text{rank}(X) \leq r. \end{aligned} \quad (3)$$

ここで $W \in \{0, 1\}^{m \times n}$ は観測に対応する重み行列であり、 $W_{ij} = 1$ ならば $V_{ij} \in \mathbb{R}$ 、そうでなければ $W_{ij} = 0$ である。 $\Omega = \{(i, j) : V_{ij} \in \mathbb{R}\} = \{(i, j) : W_{ij} = 1\}$ を用いると、問題 (3) は以下の等価な問題に定式化できる：

$$\begin{aligned} & \text{minimize} \quad \sum_{(i, j) \in \Omega} (X_{ij} - V_{ij})^2 \\ & \text{subject to} \quad \text{rank}(X) \leq r. \end{aligned} \quad (4)$$

われわれもこれと同様に Ω を用いて定式化を行なう。Olsson and Oskarsson (2009) は問題 (3) に対する凸最適化による近似に基づくヒューリスティクスを提案した。Gillis and Glineur (2011) は問題 (3) および問題 (4) の NP 困難性を証明した。

一方、行列の低ランク近似は行列のどの要素も欠損していなければ容易に行なうことができる。実際、 Ω が添え字の集合全体であるときは、問題 (4) は以下の問題を考えていることに等しい:

$$\begin{aligned} & \text{minimize} && \| \mathbf{X} - \mathbf{V} \|_{\text{F}}^2 \\ & \text{subject to} && \text{rank}(\mathbf{X}) \leq r. \end{aligned}$$

この問題は非凸であるが、 \mathbf{V} の打ち切り特異値分解によって大域的最適解を求めることができる。具体的には、この問題の最適解は $\sum_{l=1}^r \sigma_l \mathbf{p}_l \mathbf{q}_l^{\top}$ (Trefethen and Bau, 1997, Theorem 5.9) とかけることがよく知られている。ここで σ_l , \mathbf{p}_l , \mathbf{q}_l はそれぞれ \mathbf{V} の l 番目に大きい特異値とそれに対応する特異ベクトルである。 \mathbf{V} のすべての特異値と特異ベクトルを計算するためには多くの計算コストが必要である。より具体的には、 $O(\min\{m^2n, mn^2\})$ の計算量が必要であり、大きな問題例の場合は莫大な計算コストがかかる。しかし、大きい順に r 個の特異値とそれに対応する特異ベクトルは、 r が小さければ反復法により高速に計算することができる。これを打ち切り特異値分解という。

欠損値を含む行列に対して打ち切り特異値分解を用いるには、はじめに行列を補完してから特異値分解を行なう必要がある。このための方法の 1 つは、欠損が生じている要素と同じ行、あるいは列において、欠損していない入力の実数値ではじめに補完するというものである (Sarwar et al., 2000)。以下にその具体的な方法を示す:

1. 欠損のある行列 \mathbf{V} において、欠損が生じている列における観測値の平均値で、その列の欠損を一時的に埋める。この補完された行列を $\hat{\mathbf{V}}$ とする。つまり、 $(i, j) \notin \Omega$ について $\hat{V}_{ij} = (1/|\{i' : (i', j) \in \Omega\}|) \sum_{i': (i', j) \in \Omega} V_{i'j}$ である。
2. 列平均ベクトル $\boldsymbol{\mu}$ を計算する。 $\boldsymbol{\mu}$ の i 番目の要素 μ_i は \mathbf{V} における i 番目の行の観測値の平均値である。つまり、 $\mu_i = (1/|\{j : (i, j) \in \Omega\}|) \sum_{j: (i, j) \in \Omega} V_{ij}$ である。
3. 打ち切り特異値分解を用いて、 $\hat{\mathbf{V}} - \boldsymbol{\mu} \mathbf{1}^{\top}$ の最良ランク r 近似 \mathbf{V}_r を計算する。ここで $\mathbf{1}$ はすべての要素が 1 であるベクトルである。
4. 行列 \mathbf{V} の低ランク近似行列として $\mathbf{V}_r + \boldsymbol{\mu} \mathbf{1}^{\top}$ を返す。

1.3 この研究の貢献・論文の構成

この論文の構成は次の通りである: 第 2 節は行列の低ランク近似に対する新たなモデルの提案を行なう。第 3 節は提案モデルが NP 困難であることを示す。われわれが提案するモデルは問題 (4) を含むため、NP 困難である。しかしながら、われわれのモデルは上下制限約をもつため、他の先行モデルよりも協調フィルタリングに適している可能性がある。第 4 節は提案モデルに対するアルゴリズムを提案し、その収束性を示す。このアルゴリズムは交互最小化アルゴリズムであり、各反復において解く 2 つの子問題は打ち切り特異値分解と閾値処理によって解くことができる。これらは大きな問題例に対しても実用的な計算時間で実行可能である。第 5 節では予備実験の結果を示す。第 6 節では結論を述べる。

2 モデル

この節では、行列の低ランク近似に対する新たなモデルを提案する。ここで、行列 $\mathbf{X} \in \mathbb{R}^{m \times n}$ は欠損のある行列 $\mathbf{V} \in (\mathbb{R} \cup \{?\})^{m \times n}$ を補完した行列である。この \mathbf{V} と \mathbf{X} の差を最小化することを考える。この考え自体は先行研究にあるものである。

\mathbf{X} に関する制約条件として、上下限制約 $\mathbf{L} \leq \mathbf{X} \leq \mathbf{U}$ を考える。ここで $\mathbf{L} \in (\mathbb{R} \cup \{-\infty\})^{m \times n}$, $\mathbf{U} \in (\mathbb{R} \cup \{+\infty\})^{m \times n}$ は $\mathbf{L} \leq \mathbf{U}$ を満たし、不等式は要素ごとに成り立つとする。例えば、 $\mathbf{L} \leq \mathbf{X}$ はすべての i, j に対して $L_{ij} \leq X_{ij}$ であることを意味する。協調フィルタリングにおいて、ユーザはアイテムをなんらかの範囲の値で評価するため、欠損のある行列に対する低ランク近似を協調フィルタリングに応用する際にはこの値の範囲を考慮することが重要である。例えば、Amazon のユーザはアイテムを 1 から 5 までの星の数で評価する。われわれはこのような値の範囲を上下限制約としてモデル化することとし、 \mathbf{L}, \mathbf{U} をそれぞれ評価値の範囲の下限と上限を表す行列とする。この上下限は観測された要素についてはより厳しいものとしてもよい。例えば、ある要素 (i, j) に対する観測値 V_{ij} が δ 程度の誤差を含む可能性がある場合、 $L_{ij} = V_{ij} - \delta$, $U_{ij} = V_{ij} + \delta$ などとすることが考えられる。

ランク制約と上下限制約が同時に満たされないことを避けるため、各制約を個々に満たす変数を考える。このため、別の新たな変数 $\mathbf{Y} \in \mathbb{R}^{m \times n}$ を導入し、上下限制約を \mathbf{Y} に課す。さらに、目的関数に \mathbf{X} と \mathbf{Y} の差の 2 乗 Frobenius ノルム $\|\mathbf{X} - \mathbf{Y}\|_F^2$ をペナルティ項として追加する。

以上をまとめると、欠損のある行列に対する低ランク近似は以下の問題として定式化される：

$$\begin{aligned} & \text{minimize} && \|\mathbf{X} - \mathbf{Y}\|_F^2 + \lambda \sum_{(i,j) \in \Omega} (Y_{ij} - V_{ij})^2 \\ & \text{subject to} && \text{rank}(\mathbf{X}) \leq r, \\ & && \mathbf{L} \leq \mathbf{Y} \leq \mathbf{U}. \end{aligned} \tag{5}$$

ここで λ は 2 つの項のウェイトを決定するためのパラメータである。値の大きな λ を用いると、目的関数の後半部分が強調され、ランク制約に関しては多少の違反を許しつつも \mathbf{V} との差が小さい \mathbf{Y} が得られることが期待される。値の小さな λ を用いる場合は、目的関数の前半部分が強調されるため、多少 \mathbf{V} との差が大きくなるかもしれないが、2 つの制約を同時に満たす \mathbf{Y} が得られることが期待できる。

3 NP 困難性

この節では問題 (5) の NP 困難性を示す。具体的には、以下の定理を用いる：

定理 1. $\mathbf{V} \in ([0, 1] \cup \{?\})^{m \times n}$ かつ $r = 1$ のとき、問題 (5) において、目的関数値が $2^{-12}(mn)^{-7}$ より小さい精度で近似解を求めることは NP 困難である。

この定理を証明するために、以下の定理を用いる：

定理 2 (Gillis and Glineur (2011, Theorem 1.2)). $V \in ([0, 1] \cup \{?\})^{m \times n}$ かつ $r = 1$ であるとき、問題 (4) において、目的関数値が $2^{-12}(mn)^{-7}$ より小さい精度で近似解を求めることは NP 困難である。

Proof of Theorem 1. 問題 (5) において、 $(i, j) \in \Omega$ に対して $L_{ij} = U_{ij} = V_{ij}$ とし、 $(i, j) \notin \Omega$ のときは $L_{ij} = -\infty$ かつ $U_{ij} = +\infty$ とした問題例を考える。このとき、 $(i, j) \in \Omega$ について $Y_{ij} = V_{ij}$ となるため、目的関数の後半部分は 0 となって消える。結果として目的関数は $\sum_{(i,j) \in \Omega} (X_{ij} - V_{ij})^2 + \sum_{(i,j) \notin \Omega} (X_{ij} - Y_{ij})^2$ とかける。さらに、 $(i, j) \notin \Omega$ における Y_{ij} には制約がないことから最適解においては X_{ij} と一致するため、この式の後半部分は 0 となって消える。この結果、問題 (4) が得られる。この手順を踏むと、問題 (4) は問題 (5) に帰着させることができる。この帰着は明らかに多項式時間である。したがって、問題 (5) は NP 困難である。 \square

注意 1. 定理 2 は任意の r に拡張できる (Gillis and Glineur, 2011, Remark 3) ため、定理 1 は任意の r に容易に拡張することができる。

4 アルゴリズム

この節では、問題 (5) を解くための交互最小化アルゴリズムを提案する。以下ではアルゴリズムを記述するために次の関数を用いる：

$$\begin{aligned} f_0(\mathbf{X}, \mathbf{Y}) &= \|\mathbf{X} - \mathbf{Y}\|_F^2 + \lambda \sum_{(i,j) \in \Omega} (Y_{ij} - V_{ij})^2, \\ f_1(\mathbf{X}) &= \iota(\text{rank}(\mathbf{X}) \leq r), \\ f_2(\mathbf{Y}) &= \iota(L \leq \mathbf{Y} \leq U), \\ f(\mathbf{X}, \mathbf{Y}) &= f_0(\mathbf{X}, \mathbf{Y}) + f_1(\mathbf{X}) + f_2(\mathbf{Y}), \end{aligned}$$

ここで ι は指示関数である。例えば、 $\text{rank}(\mathbf{X}) \leq r$ であれば $f_1(\mathbf{X}) = 0$ ；そうでなければ $f_1(\mathbf{X}) = +\infty$ である。

$f(\mathbf{X}, \mathbf{Y})$ の最小化は問題 (5) と等価である。また、後に示すように $f(\mathbf{X}, \mathbf{Y})$ の \mathbf{X} あるいは \mathbf{Y} に関する最小化は容易に計算できる。われわれはこの事実に基づき $f(\mathbf{X}, \mathbf{Y})$ の交互最小化アルゴリズムを提案する。提案するアルゴリズムの擬似コードをアルゴリズム 1 に示す。

アルゴリズム 1 問題 (5) に対する交互最小化アルゴリズム

適当な初期値 $(\mathbf{X}^{(0)}, \mathbf{Y}^{(0)}) \in \text{dom } f_1 \times \text{dom } f_2$ をとる。

for $k = 0, 1, 2, \dots$ (収束するまで):

$\mathbf{X}^{(k+1)} = \text{argmin}_{\mathbf{X}} f(\mathbf{X}, \mathbf{Y}^{(k)})$ と更新する。

$\mathbf{Y}^{(k+1)} = \text{argmin}_{\mathbf{Y}} f(\mathbf{X}^{(k+1)}, \mathbf{Y})$ と更新する。

\mathbf{X} の更新において、 $\mathbf{X}^{(k+1)}$ は $\mathbf{Y}^{(k)}$ の最良のランク r 近似である。つまり、 $\mathbf{X}^{(k+1)}$ は以下の

子問題の最適解である:

$$\begin{aligned} & \text{minimize} \quad \|\mathbf{X} - \mathbf{Y}^{(k)}\|_F^2 \\ & \text{subject to} \quad \text{rank}(\mathbf{X}) \leq r. \end{aligned}$$

この子問題は非凸最適化問題であるが、最適解は $\mathbf{Y}^{(k)}$ の打ち切り特異値分解によって容易に計算できる。

\mathbf{Y} の更新も同様に容易に計算できる。 \mathbf{Y} を更新するには以下の子問題を解けばよい:

$$\begin{aligned} & \text{minimize} \quad \|\mathbf{X}^{(k+1)} - \mathbf{Y}\|_F^2 + \lambda \sum_{(i,j) \in \Omega} (Y_{ij} - V_{ij})^2 \\ & \text{subject to} \quad \mathbf{L} \leq \mathbf{Y} \leq \mathbf{U}. \end{aligned} \tag{6}$$

この子問題は要素ごとに解くことができる。各要素に関する子問題は、閉区間上での 1 変数の凸 2 次関数の最小化問題となる。具体的には、各 $(i, j) \in \Omega$ に対しては

$$\begin{aligned} & \text{minimize} \quad (1 + \lambda)Y_{ij}^2 - 2(X_{ij}^{(k+1)} + \lambda V_{ij})Y_{ij} \\ & \text{subject to} \quad L_{ij} \leq Y_{ij} \leq U_{ij} \end{aligned}$$

を解き、各 $(i, j) \notin \Omega$ に対しては

$$\begin{aligned} & \text{minimize} \quad Y_{ij}^2 - 2X_{ij}^{(k+1)}Y_{ij} \\ & \text{subject to} \quad L_{ij} \leq Y_{ij} \leq U_{ij} \end{aligned}$$

を解けばよい。これらの解は以下のようにかける:

$$Y_{ij}^{(k+1)} = \begin{cases} L_{ij} & (A_{ij} \leq L_{ij}), \\ A_{ij} & (L_{ij} < A_{ij} < U_{ij}), \\ U_{ij} & (U_{ij} \leq A_{ij}). \end{cases}$$

ここで

$$A_{ij} = \begin{cases} \frac{X_{ij}^{(k+1)} + \lambda V_{ij}}{1 + \lambda} & ((i, j) \in \Omega), \\ X_{ij}^{(k+1)} & ((i, j) \notin \Omega) \end{cases}$$

である。つまり、子問題 (6) は閾値処理によって解くことができ、これにかかる計算量は $O(mn)$ である。

アルゴリズム 1 によって生成された点列は、緩い仮定の下で f の停留点に収束する。ここでは、以下の 2 つの条件が成り立つならば、Tseng (2001) の意味で点 $(\bar{\mathbf{X}}, \bar{\mathbf{Y}})$ は f の停留点と呼ぶことにする。2 つの条件とは、

$$(\bar{\mathbf{X}}, \bar{\mathbf{Y}}) \in \text{dom } f = \{(\mathbf{X}, \mathbf{Y}) : f(\mathbf{X}, \mathbf{Y}) < +\infty\}$$

と

$$f'(\bar{\mathbf{X}}, \bar{\mathbf{Y}}; \Delta \mathbf{X}, \Delta \mathbf{Y}) \geq 0 \quad (\forall (\Delta \mathbf{X}, \Delta \mathbf{Y}))$$

である。ここで $f'(\bar{\mathbf{X}}, \bar{\mathbf{Y}}; \Delta \mathbf{X}, \Delta \mathbf{Y})$ は $(\bar{\mathbf{X}}, \bar{\mathbf{Y}})$ において、 $(\Delta \mathbf{X}, \Delta \mathbf{Y})$ 方向への f の下側方向微分、すなわち、

$$f'(\bar{\mathbf{X}}, \bar{\mathbf{Y}}; \Delta \mathbf{X}, \Delta \mathbf{Y}) = \liminf_{\epsilon \downarrow 0} \frac{f(\bar{\mathbf{X}} + \epsilon \Delta \mathbf{X}, \bar{\mathbf{Y}} + \epsilon \Delta \mathbf{Y}) - f(\bar{\mathbf{X}}, \bar{\mathbf{Y}})}{\epsilon}$$

である。この定義は f が不連続であっても成り立つ。具体的には、次のような定理が成り立つ。

定理 3. $\{(\mathbf{X}^{(k)}, \mathbf{Y}^{(k)})\}$ はアルゴリズム 1 によって生成された点列とし、レベル集合 $L = \{(\mathbf{X}, \mathbf{Y}) : f(\mathbf{X}, \mathbf{Y}) \leq f(\mathbf{X}^{(0)}, \mathbf{Y}^{(0)})\}$ は有界であるとする。このとき、 $\{(\mathbf{X}^{(k)}, \mathbf{Y}^{(k)})\}$ は少なくとも 1 つの集積点をもつ。さらに、すべての集積点は f の停留点である。

この定理を証明するため、以下の補題を用いる。

補題 1. $\text{dom } f$ は閉集合である。

Proof. 明らかに、 $\text{dom } f = \text{dom } f_1 \times \text{dom } f_2$ であり $\text{dom } f_2$ は有界である。したがって、 $\text{dom } f_1$ の有界性のみを示せばよい。 $\mathbf{X} \notin \text{dom } f_1$ を任意にとる。すると、 r 番目に大きな \mathbf{X} の特異値 $\sigma_r(\mathbf{X})$ は正である。そして $\|\mathbf{E}\|_2 < \sigma_r(\mathbf{X})$ となるような \mathbf{E} を任意にとる。Golub and van Loan (2013, Corollary 8.6.2) より、 $\mathbf{X} + \mathbf{E} \notin \text{dom } f_1$ であるために、 $l = 1, \dots, r$ に対して $\sigma_l(\mathbf{X} + \mathbf{E}) \geq \sigma_l(\mathbf{X}) - \|\mathbf{E}\|_2 > 0$ が容易に示せる。これは $\text{dom } f_1$ が閉であることを意味する。□

この補題を用いて定理 3 を示す。この証明は本質的に Tseng (2001, Sections 3 and 4) に基づく。

定理 3 の証明。各更新における最適性から、点列 $\{(\mathbf{X}^{(k)}, \mathbf{Y}^{(k)})\}$ は $L \subset \text{dom } f$ に含まれる。したがって、点列 $\{(\mathbf{X}^{(k)}, \mathbf{Y}^{(k)})\}$ は有界。補題 1 より $\text{dom } f$ が閉であるため、 $\{(\mathbf{X}^{(k)}, \mathbf{Y}^{(k)})\}$ は $\text{dom } f$ において少なくとも 1 つの集積点をもつ。

$\{(\mathbf{X}^{(k_j)}, \mathbf{Y}^{(k_j)})\}$ を、集積点 $(\bar{\mathbf{X}}, \bar{\mathbf{Y}})$ に収束する $\{(\mathbf{X}^{(k)}, \mathbf{Y}^{(k)})\}$ の部分列とする。 $(\bar{\mathbf{X}}, \bar{\mathbf{Y}})$ が停留点であるということを示すには、 $(\bar{\mathbf{X}}, \bar{\mathbf{Y}}) \in \text{dom } f$ であるため、任意の $(\Delta \mathbf{X}, \Delta \mathbf{Y})$ において $f'(\bar{\mathbf{X}}, \bar{\mathbf{Y}}; \Delta \mathbf{X}, \Delta \mathbf{Y}) \geq 0$ となることを示せばよい。各更新における最適性から、

$$f(\mathbf{X}^{(k_{j+1})}, \mathbf{Y}^{(k_{j+1})}) \leq f(\mathbf{X}^{(k_{j+1})}, \mathbf{Y}^{(k_j)}) \leq f(\mathbf{X}, \mathbf{Y}^{(k_j)}) \quad (\forall \mathbf{X})$$

を得る。 f は $\text{dom } f$ において連続であるため、 $j \rightarrow \infty$ の極限を取ると

$$f(\bar{\mathbf{X}}, \bar{\mathbf{Y}}) \leq f(\mathbf{X}, \bar{\mathbf{Y}}) \quad (\forall \mathbf{X}) \quad (7)$$

を得る。さらに、

$$f(\mathbf{X}^{(k_j)}, \mathbf{Y}^{(k_j)}) \leq f(\mathbf{X}^{(k_j)}, \mathbf{Y}) \quad (\forall \mathbf{Y})$$

について同様に $j \rightarrow \infty$ の極限を取ると

$$f(\bar{\mathbf{X}}, \bar{\mathbf{Y}}) \leq f(\bar{\mathbf{X}}, \mathbf{Y}) \quad (\forall \mathbf{Y}) \quad (8)$$

を得る。 f_0 は微分可能であるので、任意の $(\Delta \mathbf{X}, \Delta \mathbf{Y})$ について以下の不等式が成り立つ：

$$\begin{aligned} & f'(\bar{\mathbf{X}}, \bar{\mathbf{Y}}; \Delta \mathbf{X}, \Delta \mathbf{Y}) \\ &= \langle \nabla f_0(\bar{\mathbf{X}}, \bar{\mathbf{Y}}), (\Delta \mathbf{X}, \Delta \mathbf{Y}) \rangle + \liminf_{\epsilon \downarrow 0} \left(\frac{f_1(\bar{\mathbf{X}} + \epsilon \Delta \mathbf{X}) - f_1(\bar{\mathbf{X}})}{\epsilon} + \frac{f_2(\bar{\mathbf{Y}} + \epsilon \Delta \mathbf{Y}) - f_2(\bar{\mathbf{Y}})}{\epsilon} \right) \end{aligned}$$

$$\begin{aligned}
&\geq \langle \nabla_{\mathbf{X}} f_0(\bar{\mathbf{X}}, \bar{\mathbf{Y}}), \Delta \mathbf{X} \rangle + \langle \nabla_{\mathbf{Y}} f_0(\bar{\mathbf{X}}, \bar{\mathbf{Y}}), \Delta \mathbf{Y} \rangle \\
&\quad + \liminf_{\epsilon \downarrow 0} \frac{f_1(\bar{\mathbf{X}} + \epsilon \Delta \mathbf{X}) - f_1(\bar{\mathbf{X}})}{\epsilon} + \liminf_{\epsilon \downarrow 0} \frac{f_2(\bar{\mathbf{Y}} + \epsilon \Delta \mathbf{Y}) - f_2(\bar{\mathbf{Y}})}{\epsilon} \\
&= \langle \nabla_{\mathbf{X}} f_0(\bar{\mathbf{X}}, \bar{\mathbf{Y}}), \Delta \mathbf{X} \rangle + \langle \nabla_{\mathbf{Y}} f_0(\bar{\mathbf{X}}, \bar{\mathbf{Y}}), \Delta \mathbf{Y} \rangle + f'_1(\bar{\mathbf{X}}; \Delta \mathbf{X}) + f'_2(\bar{\mathbf{Y}}; \Delta \mathbf{Y}) \\
&= \liminf_{\epsilon \downarrow 0} \frac{f(\bar{\mathbf{X}} + \epsilon \Delta \mathbf{X}, \bar{\mathbf{Y}}) - f(\bar{\mathbf{X}}, \bar{\mathbf{Y}})}{\epsilon} + \liminf_{\epsilon \downarrow 0} \frac{f(\bar{\mathbf{X}}, \bar{\mathbf{Y}} + \epsilon \Delta \mathbf{Y}) - f(\bar{\mathbf{X}}, \bar{\mathbf{Y}})}{\epsilon} \\
&\geq 0.
\end{aligned}$$

ここで、最後の不等式は式 (7) と (8) より成り立つ。 \square

注意 2. レベル集合 $L = \{(\mathbf{X}, \mathbf{Y}) : f(\mathbf{X}, \mathbf{Y}) \leq f(\mathbf{X}^{(0)}, \mathbf{Y}^{(0)})\}$ の有界性は、例えば、任意の i, j について $L_{ij} > -\infty$ かつ $U_{ij} < +\infty$ であるときに成り立つ。

注意 3. われわれの提案アルゴリズムは、 \mathbf{Y} に対する上下限制約の代わりに閉凸制約を用いるものに対しても拡張できる。 \mathbf{Y} を更新するための子問題が効率的に解ければ、アルゴリズムは効率よく動く。実際には、 \mathbf{Y} の打ち切り特異値分解よりも高速に \mathbf{Y} を更新するための子問題を解くことができれば、各反復における全体の計算量は大きく変わらない。

5 数値実験

この節では、提案アルゴリズムの収束の速さや収束先に初期点の選択がどのような影響を及ぼすかについて調べた計算機実験の結果を示す。この実験に用いた計算機の OS は MacOS Sierra 10.12.6, CPU は Intel Core m3 1.1 GHz, RAM は 8 GB である。アルゴリズムは MATLAB (R2017b) で実装した。

この実験では人工的に問題例を生成した。まず、 $\text{rank}(\mathbf{A}) = 10$ かつすべての i, j において $0.5 \leq A_{ij} \leq 5.5$ であるような行列 $\mathbf{A} \in \mathbb{R}^{20 \times 100}$ を次の手順で生成した：まず、2 つの行列 $\mathbf{B} \in \mathbb{R}^{20 \times 9}$, $\mathbf{C} \in \mathbb{R}^{9 \times 100}$ の各成分を標準正規分布から独立にサンプリングした。次に、これらの積に対して $0.5 \leq A_{ij} \leq 5.5$ となるようにスケーリングを行なった。続いて \mathbf{A} の要素のすべてを $\bar{A}_{ij} \in \{1, 2, \dots, 5\}$ と整数に丸め、要素のうち 80% を欠損させた。この欠損のある行列を \mathbf{V} として用いた。われわれのアルゴリズムでは、打ち切り特異値分解を行なうときのパラメータ r を $r = 10$ とし、目的関数における重みパラメータ λ を $\lambda = 1$ とした。

この実験では、初期点の生成法に関して 4 つの方法を比較した。

SKKR \mathbf{V} に対して Sarwar et al. (2000) によって提案されたアルゴリズムを適用し、得られた解をわれわれの提案アルゴリズムの初期点として用いる。

perturb + SKKR 摂動させた \mathbf{V} に対して Sarwar et al. (2000) によって提案されたアルゴリズムを適用し、得られた解をわれわれの提案アルゴリズムの初期点として用いる。

low rank rand \mathbf{A} を生成するときと同様の手順で初期点を生成する。

rand 20×100 行列の要素を標準正規分布から独立にサンプリングし、初期点として用いる。

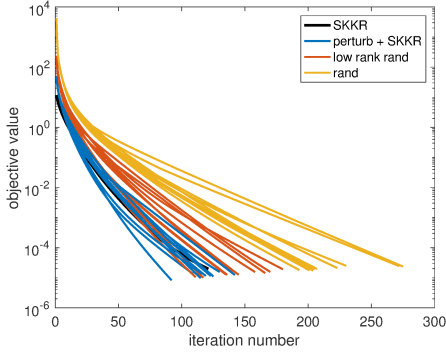


図 1: 反復回数の比較

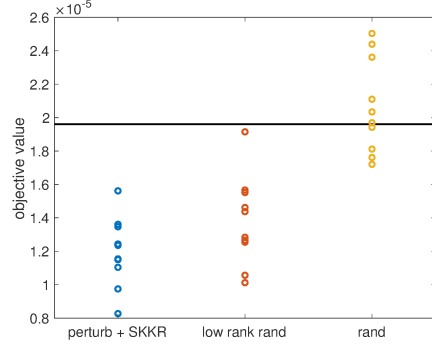


図 2: 目的関数値の比較

SKKR については 1 個, perturb + SKKR, low rank rand, and rand のそれぞれについて 10 個の初期点を生成し, 提案アルゴリズムを用いて異なる解を得た。

停留点に収束するために必要な反復回数の違いを図 1 に示す。図の横軸は反復回数, 縦軸は目的関数値である。図 1 より, perturb + SKKR と low rank rand によって生成された初期点を用いると, SKKR によって生成された初期点を用いたときよりも速く収束したことがわかる。一方, rand によって生成された初期点を用いると収束に時間がかかったことがわかる。

停留点における目的関数値を図 2 に示す。縦軸は目的関数値を表す。なお, 2×10^{-5} 付近にある黒い横線は, SKKR によって得た初期点を提案アルゴリズムに適用したときの収束先における目的関数値である。図 2 から, perturb + SKKR と low rank rand によって生成された初期点を用いたときのすべての解は, SKKR によって生成された初期点から得られた解よりも目的関数値が小さかったことがわかる。rand から得られた初期点のうちのいくつかは同様に, SKKR によって得られた初期点を用いて得られた解より目的関数値が小さかった。

6 結論

本論文では, 欠損のある行列を補完するための低ランク近似モデルを提案した。提案する最適化モデルはランク制約だけでなく上下限制約をもつ。上下限制約を考慮しているため, 提案モデルは推薦システムに有効である可能性がある。さらに, 提案モデルに対する交互最小化アルゴリズムを提案し, 緩い仮定の下で提案されたアルゴリズムによって生成される点列が停留点に収束することを示した。数値実験の結果は予備的なものではあるが, 提案アルゴリズムの性能は初期点に依存することがわかった。そのため, 実用的には多数の初期点を試し, 最も良い解を選ぶ必要があると考えられる。また, 本論文における実験は, パラメータ r と λ を固定して行なったが, 実用的には交差検証などにより適当なパラメータを選択する必要があるだろう。より詳細な数値実験をすることは今後の課題とする。

REFERENCES

- Candés, E. J. and Recht, B. (2009). Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772.
- Gillis, N. and Glineur, F. (2011). Low-rank matrix approximation with weights or missing data is NP-hard. *SIAM Journal on Matrix Analysis and Applications*, 32(4):1149–1165.
- Golub, G. H. and van Loan, C. F. (2013). *Matrix Computations*. The Johns Hopkins University Press, fourth edition.
- Ilin, A. and Raiko, T. (2010). Practical approaches to principal component analysis in the presence of missing values. *Journal of Machine Learning Research*, 11(Jul):1957–2000.
- Olsson, C. and Oskarsson, M. (2009). A convex approach to low rank matrix approximation with missing data. In Salberg, A.-B., Hardeberg, J. Y., and Jenssen, R., editors, *Proceedings of the 16th Scandinavian Conference on Image Analysis (SCIA '09)*, pages 301–309.
- Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2000). Application of dimensionality reduction in recommender system—a case study. In Kohavi, R., Masand, B., Spiliopoulou, M., and Srivastava, J., editors, *Proceedings of the ACM WEBKDD 2000 Workshop*.
- Su, X. and Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 2009(421425):1–19.
- Tipping, M. E. and Bishop, C. M. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society. Series B*, 61(3):611–622.
- Trefethen, L. N. and Bau, D. (1997). *Numerical Linear Algebra*. SIAM.
- Tseng, P. (2001). Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of Optimization Theory and Applications*, 109(3):475–494.